

Lucia Ku  
Professor Avinash Jairam  
CIS 3120  
2 April 2022

### Project 1 Report

In this project, our goal is to create a program using the Python programming language in order to compare and contrast the various heights of the different athletes in 40 different college sports teams. We are provided with 40 different web pages consisting of a roster full of names and heights for each team.

In the first block of code, my first step was to import the different libraries needed to run the program. I imported beautiful soup, requests, pandas, and files. In the second block of code, I created dictionaries for each sport using the 40 web page links provided, with the college name as a key and the URL as a value. As a result, four dictionaries were created: "mens\_volleyball," "womens\_volleyball," "mens\_swimming," and "womens\_swimming," each with their own individual colleges and roster links. In the third block of code, I created a list called "sportsteam" that contained these four dictionaries so I would be able to reference the list in a later step and loop through each dictionary. In the fourth block of code, I created a function called "scrape" in order to scrape information from the web pages. Underneath the scrape function, I created two empty lists, one for height and the other for names. After that, I created a for loop in order to continuously loop through each URL in the dictionaries. Inside the for loop, I first defined "page" to get the information listed in each URL. Then, I used beautiful soup and "page" to parse the information that was gathered from each URL, and defined that as "soup." Using "soup," I created "all\_relevant\_name" to specifically locate "sidearm-table-player-name" in the HTML code, which will bring back the names of all the athletes, and "all\_relevant\_height" to specifically locate "height" in the HTML code, which will bring back the heights of all the athletes. Next, I created another for loop inside this for loop in order to loop through the "all\_relevant\_height" dataset. I used an if function to add in a "-" to any empty heights in the height list with an append function. Otherwise, the if function will take the existing height and first split the feet from the inches into separate lists and then convert the height into inches by multiplying the feet by 12 and adding it to the inches. Next, I move out of the inner for loop but while still staying inside of the outer for loop, I first set "total" to 0. Then, I create another for loop inside to add any heights without a "-" to the "total" with an if function. After that, I move out of the inner for loop to set "height\_average" equal to the "total" divided by the number of athletes, which will result in the average height of athletes. Next, I create another inner for loop in order to set the empty heights represented by "-" equal to the average heights. By doing so, we will be able to fill in any missing height data by replacing them with the collective average height of the team so as to not offset the current information gathered. After moving out of that inner for loop, I created another inner for loop for the "all\_relevant\_name" dataset in order to get rid of any leading and trailing spaces with the "strip" function and added it to the "name" list with the "append" function. After that, I moved out of the outer for loop and set "height\_average" equal to the "total" divided by the number of athletes again and printed out the result afterwards. Next, I set "filename" equal to the dictionary file name plus ".csv" at the end in order to name what the file will be when downloaded as a csv file. Then, I created a dataframe and named it "df." After that, I added the name and height lists to the dataframe, and printed the dataframe after. Next, I sorted the players by their height in descending order using the dataframe, and set that step equal to "tallest." In order to print the result accordingly, in the next line, I used the ".head" and ".tail" functions to specifically print out the top 5 tallest and top 5 shortest players in each sport. For the shortest players, I also called the "tallest" step but sorted the height in ascending order instead. After that, I converted the

dataframe into a csv file and downloaded it. Lastly, I returned the “height\_average,” which will display the average height of athletes in each sport. All of the information in the fourth block of code was defined under the function “scrape.” In the fifth block of code, I created an empty list called “averages” to store average heights in, and an empty list called “names” to store athlete names in. Then, I created a for loop in order to loop through the list of dictionaries called “sportsteam” that was created in the third block of code. Inside that for loop, I created another for loop in order to loop through each dictionary. Under that, I formatted the result by separating the sections by sports name so as to be able to read the resulting data clearer. I also added the height averages of each dictionary to the “averages” list and the names of each athlete into the “names” list using the “append” function. After that, I moved out of both for loops and printed the “averages” list. Then, I created another dataframe called “average\_player\_height” to display the athlete names and height averages collected within the “averages” and “names” lists. In the final block of code, I created a bar graph using “average\_player\_height” while attaching “.plot.bar” to the end. I named the x axis “sport,” named the y axis “average,” rotated the y axis names by “10” for formatting purposes, and named the bar graph “Average Heights of Athletes.”

According to the information gathered from the program, the average height for the men’s volleyball teams is 73.04 inches, the average height for the women’s volleyball teams is 67.41 inches, the average height for the men’s swimming teams is 71.53 inches, and the average height for the women’s swimming teams is 65.10 inches. As a result, if we examine the information gathered and the bar graph, we can conclude that on average, volleyball players are slightly taller than swimmers and male athletes are generally taller than female athletes. If compared to the results from Homework 1, we can see that although the average numbers differ slightly, the results generally stayed the same when comparing heights between volleyball players and swimmers, and likewise, male and female athletes. Therefore, the program that I created for this project can definitely be representative of both sports in regards to college volleyball players and swimmers in New York, since we have a large sample and proven data trends.

